Conciliación de la población de los censos y las estadísticas de nacimientos, defunciones y migración a través de una función expolinomial

Manuel Ordorica Mellado*

El objetivo del artículo es la construcción de una función matemática que describa la dinámica de la población total en México entre 1940 y 1990. A dicha función se le ha llamado expolinomial. En el trabajo se intenta, además, realizar un análisis de conciliación de los datos de población producidos a partir de los censos con la información de nacimientos, defunciones y migración.

Introducción

El objetivo del presente artículo es la construcción de una función matemática que describa la evolución de la población total de México entre 1940 y 1990. Se intenta además realizar un análisis de conciliación de los datos de población producidos a partir de los censos, con la información de nacimientos, defunciones y migración. El análisis sólo es válido para el periodo 1940-1990.

Los datos de la población total son ajustados, la mayoría de las veces, mediante funciones matemáticas, sin que se tome en consideración la dinámica de los componentes del crecimiento demográfico, tal es el caso de la exponencial y la logística, entre otras.

Por tanto, es importante desarrollar un modelo que vincule las transformaciones de los niveles de natalidad, de mortalidad y de migración con la dinámica del número de personas.

A partir de los cambios en los componentes se busca desarrollar un modelo que genere las cifras de población, tomando en cuenta sólo el dato inicial o de partida. Esto significa que, si disponemos de las tasas de natalidad, de mortalidad y de migración para el periodo en estudio, y la población al inicio del mismo, es posible obtener el número de habitantes para todos los años. El modelo que se describirá a continuación sólo considera datos globales, pero podrá ser aplicado a datos por grupos de edades y sexo.

El periodo elegido para la aplicación del modelo es de 1940 a 1990, esto es, 50 años de evolución demográfica.

*Profesor-investigador del Centro de Estudios Demográficos y de Desarrollo Urbano de El Colegio de México.

Fuentes de información y metodología

Se han utilizado las tasas de natalidad, mortalidad y migración para el periodo de 1940 a 1950 y la población censal de 1940 proyectada al 30 de junio.

Sea

$$r(t) = \frac{1}{P(t)} \frac{\partial P(t)}{\partial t}$$

$$= b(t) - d(t) + i(t) - e(t) \text{ para toda } t$$

donde $r(t)$ es la tasa de crecimiento de la población en t
$P(t)$ es la población en el momento t
$b(t)$ es la tasa de natalidad en t
$d(t)$ es la tasa de mortalidad en t
$i(t)$ es la tasa de inmigración en t
$e(t)$ es la tasa de emigración en t

Si bien es cierto que la inmigración se ha incrementado, su número no incide de manera significativa sobre el ritmo de crecimiento demográfico. Por tanto, se supone que $i(t) = 0$, para toda t.

La tasa de crecimiento demográfico es igual a la tasa de natalidad menos la tasa de mortalidad menos la tasa de emigración. En consecuencia, se puede escribir la siguiente relación:

$$r(t) = b(t) - d(t) - e(t)$$

Supongamos que la evolución de cada uno de los componentes del crecimiento demográfico puede describirse a partir de un polinomio, para lo cual se utilizará la técnica de métodos numéricos. Se sabe que es posible reproducir una serie de datos mediante el ajuste de un polinomio.

Sean

$$b(t) = b_0 + b_1 t + \dots + b_{n-1} t^{n-1} + b_n t^n$$

$$d(t) = d_0 + d_1 t + \dots + d_{m-1} t^{m-1} + d_m t^m$$

y

$$e(t) = e_0 + e_1 t + \dots + e_{p-1} t^{p-1} + e_p t^p$$

$$t \varepsilon [0, 50]$$
$$n, m, p \varepsilon Z$$

donde los coeficientes del polinomio son parámetros a estimar.

Se hace una transformación: a 1940 se le hace corresponder t igual a 0 y a 1990 se le hace corresponder t igual a 50. Las funciones construidas de esta manera pasan por las tasas estimadas.

La tasa de crecimiento demográfico $r(t)$ va a estar representada por un polinomio. El orden del polinomio $r(t)$ está dado por:

$$max\,\{n,m,p\}$$

Si se tiene

$$\rightarrow \frac{1}{P(t)} \; \frac{\partial P(t)}{\partial t} = r(t)$$

$$\rightarrow P(t) = P(0)\, e^{\int r(t)\, dt}$$

Esta última fórmula se obtiene resolviendo la ecuación diferencial. El exponente del número e es la integral de un polinomio, fácil de desarrollar.

Los datos utilizados se presentan a continuación:

CUADRO 1

Tasas de natalidad, mortalidad y emigración 1940-1990 (por mil)

Tasas de natalidad		
(periodo)	b(t)	
1940	44.30	(1)
1945 - 1949	45.12	(2)
1950 - 1954	46.30	(3)
1955 - 1959	45.60	(2)
1960 - 1964	44.90	(3)
1965 - 1969	44.20	(3)
1970 - 1974	42.70	(3)
1975 - 1979	37.60	(3)
1980 - 1984	32.90	(2)
1985 - 1989	30.20	(2)
1990	28.60	(2)

Tasas de mortalidad		
t (periodo)	d(t)	
1940	22.60	(2)
1945 - 1949	22.00	(4)
1950 - 1954	19.10	(2)
1955 - 1959	16.20	(3)
1960 - 1964	13.75	(2)
1965 - 1969	11.30	(3)
1970 - 1974	10.20	(3)
1975 - 1979	9.20	(3)
1980 - 1984	7.90	(3)
1985 - 1989	6.50	(5)
1990	6.00	(2)

Tasas de emigración		
t (periodo)	e(t)	
1940	0.30	(2)
1945	0.30	(6)
1950	0.35	(2)
1955	0.40	(6)
1960	0.60	(2)
1965	0.80	(6)
1970	1.90	(2)
1975	3.00	(2)
1980	2.75	(2)
1985	2.50	(2)
1990	2.50	(2)

Fuentes: (1) El Colegio de México, *Dinámica de la población de México*, CEED, 1970, p. 47.

(2) Estimaciones propias.

(3) SPP, Conapo, Celade, México, *Estimaciones-proyecciones de población, 1950-2000*, México, 1983, p. 11.

(4) El Colegio de México, *op. cit.*, p. 14.

(5) Naciones Unidas, *Perspectivas de la población mundial. Estimaciones y proyecciones en 1982*, Nueva York, UNAM, 1985, p. 353.

(6) *Revista Mexicana de Sociología*, Instituto de Investigaciones Sociales en: Una función expologística para el análisis de las fuentes demográficas entre 1940 y 1990, México, año LV, núm. 1, 1193, p. 6.

Las ecuaciones encontradas son las siguientes:

Para la tasa de natalidad:

$$b(t) = (5.375e-16)\,t^{10} - (1.254e-13)\,t^{9} + (1.219e-11)\,t^{8} - (6.328e-10)\,t^{7} + (1.889e-8)\,t^{6} - (3.244e-7)\,t^{5} + (3.002e-6)\,t^{4} - (1.272e-5)\,t^{3} + (1.040e-5)\,t^{2} + (1.847e-4)\,t + (4.43e-2)$$

Para la tasa de mortalidad:

$$d\ (t) = (6.236e - 16)\ t^{10} - (1.495e - 13)\ t^{9} + (1.517e - 11)\ t^{8} - (8.482 - e10)\ t^{7} + (2.853e - 8)\ t^{6} - (5.914e - 7)\ t^{5} + (7.413e - 6)\ t^{4} - (5.194e - 5)\ t^{3} + (1.606e - 4)\ t^{2} - (5.485e - 4)\ t + (2.26e - 2)$$

Para la tasa de emigración:

$$e\ (t) = (8.381e - 16)\ t^{10} - (2.242e - 13)\ t^{9} + (2.556e - 11)\ t^{8} - (1.620e - 9)\ t^{7} + (6.251e - 8)\ t^{6} - (1.511e - 6)\ t^{5} + (2.269e - 5)\ t^{4} - (2.020e - 4)\ t^{3} + (9.553e - 4)\ t^{2} - (1.789e - 3)\ t + (3.003e-4)$$

Los datos observados y los estimados son iguales por construcción del modelo.

Haciendo

$$F\ (t) = \int\ r\ (t)\ dt = \int\ \{b\ (t) - d\ (t) - e\ (t)\ \}\ dt$$

$$F\ (t) = -\ (8.402e - 17)\ t^{11} + (2.479e - 14)\ t^{10} - (3.170e - 12)\ t^{9} + (2.294e-10)\ t^{8} - (1.031e - 8)\ t^{7} + (2.963e - 7)\ t^{6} - (5.420e - 6)\ t^{5} + (6.032 - e5)\ t^{4} - (3.685e - 4)\ t^{3} + (1.261e - 3)\ t^{2} + (2.14e - 2)\ t$$

<u>Nota:</u> $e - x = 10^{-x}$

El grado de $F\ (t)$ es igual a: $max\ (n, m, p) + 1$

Sea PC (0) igual a 19.8 millones, que es la población de 1940, esto es, el punto de partida. Se supondrán diferentes niveles de omisión de la población del censo de 1940, a fin de obtener diferentes trayectorias. Los resultados del modelo son:

CUADRO 2
Población estimada (en millones)

Omisión %	*1940*	*1950*	*1960*	*1970*	*1980*	*1990*
0.0	19.8	25.3229	34.3037	47.5497	63.2604	79.1649
1.0	20.0	25.5787	34.6502	48.0300	63.8994	79.9646
2.0	20.2	25.8345	34.9967	48.5103	64.5384	80.7642
3.0	20.4	26.0903	35.3432	48.9906	65.1774	81.5638
4.0	20.6	26.3460	35.6897	49.4709	65.8164	82.3635
5.0	20.8	26.6018	36.0362	49.9512	66.4554	83.1631
6.0	21.0	26.8576	36.3827	50.4315	67.0943	83.9628
7.0	21.2	27.1134	36.7292	50.9118	67.7333	84.7624
8.0	21.4	27.3692	37.0757	51.3921	68.3723	85.5621
9.0	21.6	27.6250	37.4222	51.8724	69.0113	86.3617

A fin de analizar las diferentes trayectorias, se han calculado las diferencias entre la población censal y la población estimada para los diferentes valores de omisión del censo de 1940. Si bien es cierto que lo antes mencionado puede aproximarnos a la elección de una trayectoria, no existe un criterio objetivo para tomar una decisión.

CUADRO 3
Población censal menos población estimada (en millones)

Omisión %	*1940*	*1950*	*1960*	*1970*	*1980*	*1990*
0	0	0.5	0.7	1.4	3.7	1.5
1	-0.2	0.2	0.3	0.9	3.1	1.7
2	-0.4	0.0	0.0	0.4	2.5	0.9
3	-0.6	-0.3	-0.3	-0.1	1.8	0.0
4	-0.8	-0.6	-0.7	-0.6	1.2	-0.7
5	-1.0	-0.8	-1.0	-1.1	0.6	-1.5
6	-1.2	-1.1	-1.4	-1.5	-0.1	-2.3
7	-1.4	-1.3	-1.7	-2.0	-0.7	-3.1
8	-1.6	-1.6	-2.1	-2.5	-1.4	-3.9

El cuadro anterior muestra que no es posible aceptar un porcentaje de omisión del censo de 1940 menor a 3%, debido a que la población censal de 1990 es mayor a la población estimada con el modelo para este mismo año.

Criterio de selección de trayectorias

El análisis anterior no permite elegir una trayectoria específica. A fin de encontrar una trayectoria óptima se realizará el siguiente procedimiento:
Sea

$$z = \sum \{PE\,(t, k) - PC\,(t)\,\}^2$$

con $\quad$ $\mathrm{PE}\,(\mathrm{t}, \mathrm{k}) > 0$
$\mathrm{PC}\,(\mathrm{t}) > 0 \quad$ y

$$PE\,(t, k) = PC\,(0)\ \ (1{+}k)\ \ \mathrm{e}^{\int r\,(t)\,dt}$$

donde $PE\,(t, k)$ es la población estimada con el modelo y $PC\,(t)$ es la población censada proyectada al 30 de junio del año t y k es el grado de omisión de la población del censo de 1940.

El propósito es minimizar z respecto a k. El criterio es minimizar z, que es igual a la suma de cuadrados de las desviaciones entre la población censal proyectada a mitad del año t y la población estimada con el modelo para ese mismo momento.

$$\text{min. } z = \min \sum \{PE(t, k) - PC\ (t)\}^2$$

lo anterior es equivalente a:

$$\text{min. } z = \min \sum \{PC\ (0)\ (1 + k)\ e^{\int r(t)\,dt} - PC\ (t)\}^2$$

Al derivar respecto a k e igualando a cero para encontrar la condición de primer orden, se obtiene lo siguiente:

$$\frac{\partial z}{\partial k} = 2 \sum \{PC\ (0)\ (1 + k)\ \mathrm{e}^{\int r(t)\,dt} - PC\ (t)\ \}\ PC\ (0)\ e^{\int r(t)\,dt} = 0$$

de lo anterior se sigue que:

$$k = \frac{\sum PC(t)}{PC(0) \sum e^{\int r(t)\,dt}}$$

Al hacer los cálculos se obtiene que k es igual a 3.3% de omisión. A fin de analizar si el óptimo encontrado es un máximo o un mínimo, se obtiene la condición de segundo orden, que es la siguiente:

$$\frac{\partial^2 z}{\partial k^2} = 2\ \mathrm{P}\ (0)^{\,2}\ e^{\int r(t)\,dt} \sum \mathrm{e}^{\int r(t)\,dt} > 0$$

Por lo tanto, el óptimo de z es un mínimo. Es decir, el valor encontrado de k minimiza la función z. A continuación se presentan los datos estimados para $k = 3.3\%$ de omisión.

CUADRO 4
Población estimada (en millones)

Año	*Habitantes*
1940	20.5
1950	26.2
1960	35.4
1970	49.1
1980	65.4
1990	81.8

La población censal proyectada al 30 de junio se presenta a continuación:

CUADRO 5
Población censal (en millones)

Año	*Habitantes*
1940	19.8
1950	25.8
1960	35.0
1970	48.9
1980	67.0
1990	81.7

Fuente: Censos Generales de Población y Vivienda, 1940-1990.
Para la población estimada menos la población censal véanse los cuadros 6 y 7.

CUADRO 6
Población estimada - población censal (en millones)

Año	*Diferencia*
1940	0.7
1950	0.4
1960	0.4
1970	0.2
1980	-1.6
1990	0.1

CUADRO 7
(Población estimada menos población censal) / población estimada (en porcentajes)

Año	%
1940	3.3
1950	1.5
1960	1.1
1970	0.4
1980	-2.4
1990	0.1

Es difícil aceptar un grado de omisión del censo de 1990 de .1%, lo que significa que las tasas utilizadas podrían ser diferentes. Podría ocurrir, por ejemplo, que las tasas de natalidad fueran ligeramente mayores o las de emigración más bajas. También podría ocurrir que la población inicial, la de 1940, tuviera un grado mayor de omisión.

¿Qué nivel de omisión de la población del censo de 1940 podría considerarse como aceptable?

No hay investigaciones que muestren el grado de omisión de la población del censo de 1940, pero sí los hay que lo calculan para 1950. De los trabajos realizados entre 1954 y 1960 (cercanos al censo de 1950), que tomaré para dar una cota superior, se encuentran los siguientes:

La División de Población de las Naciones Unidas publicó en 1954 proyecciones de población por sexo y grupos de edades para el periodo 1950-1980 correspondientes a Centroamérica y México. La población corregida para 1950 fue de 25.6 millones de personas, cifra inferior a la población censal proyectada al 30 de junio de 1950, que fue de 25.8 millones de habitantes.

En 1958, Coale y Hoover realizaron proyecciones de población con base en la que existía en 1950. La población corregida para 1950 por los autores fue de 26.6 millones de personas, lo que significa un grado de omisión de 3.0%. Esta cifra equivale a un nivel de omisión de 5% de la población censal de 1940.

En 1960, Louis Ducoff preparó para la CEPAL un estudio sobre los problemas demográficos de Centroamérica y México y sus efectos en el desarrollo. Para ello realizó proyecciones de población para el periodo 1950-1980. La población corregida para 1950 es de 25.8 millones de habitantes; esto supone que sólo se proyectó la población a mitad de año.

La población corregida para 1950 según las proyecciones realizadas en 1954 y en 1960 no son aceptables a partir del modelo, ya que darán cifras para 1990 inferiores a las censales. Por tanto, utilizaré las cifras de 1950 estimadas por Coale y Hoover.

La trayectoria que se define a partir del dato estimado por Coale y Hoover para 1950 se presenta en el cuadro siguiente:

CUADRO 8
Población estimada suponiendo un grado de omisión de la población del censo de 1950 de 3% (en millones)

Año	*Habitantes*
1940	20.8
1950	26.6
1960	36.0
1970	49.9
1980	66.4
1990	83.2

Porcentaje de omisión de los censos

Año	%
1940	4.8
1950	3.0
1960	2.8
1970	2.0
1980	-0.9
1990	1.8

Este resultado muestra que el grado de omisión pasó de casi 5% en 1940 a 1.8%; se observa que el peor censo en cobertura es el de 1940, y si bien es cierto que los censos han mejorado la cobertura del censo de 1950, 1960 es similar. Lo mismo ocurre entre los censos de 1970 y 1990.

Conclusiones

a) Los resultados muestran que los censos han mejorado en cobertura con el transcurrir de los años, de acuerdo con las tasas de los componentes demográficos que se han usado.

b) Parecería que el de 1990 es el mejor entre los seis censos.

c) El grado de cobertura del censo de 1990 es ligeramente superior al de 1970.

d) Los datos estimados por el modelo para 1980 muestran que el censo levantado en ese año tiene una población mayor a la que se obtendría de acuerdo con las tasas utilizadas.

e) Si bien es cierto que resulta difícil aceptar un grado de omisión tan pequeño para el censo de 1990, podría decirse que éste está mejor cubierto que el de 1940 en 3.3% (3.4% - .1%). Lo mismo se observa cuando se usan los datos estimados por Coale y Hoover sobre la población de 1950. Esto significa que el modelo estima coberturas relativas entre los censos.

f) Una desventaja del modelo es que los resultados son válidos únicamente para el periodo de 1940 a 1990. No es posible hacer proyecciones de población con el modelo. Sólo sirve para hacer una conciliación entre las cifras de los componentes de la dinámica demográfica y la población de los censos en este periodo.

g) El modelo utilizado es de carácter determinístico. Será importante incorporarle un error de tipo aleatorio a fin de que proporcione bandas de confianza de las estimaciones.

h) Se supone que las tasas de natalidad, de mortalidad y de migración son confiables y precisas.

i) Los resultados de la aplicación del modelo dependen de las estimaciones de las tasas de los componentes demográficos que se utilizan. Dichas tasas se han calculado a partir de la información censal, por lo que incorporan de forma implícita las poblaciones censales. Sin embargo, el modelo debe utilizarse como una referencia que permita conocer la congruencia de las cifras demográficas.

Gran parte de las técnicas de conciliación utilizan algún dato demográfico que es considerado como pivote. En la demografía nos encontramos en un círculo vicioso: nos apoyamos en datos de dudosa confiabilidad para estimar otros índices.

Bibliografía

Keyfitz, Nathan (1979), *Introducción a las matemáticas de la población*, Santiago de Chile, Centro Latinoamericano de Demografía.

Lotka, Alfred (1969), *Teoría analítica de las asociaciones biológicas*, Santiago de Chile, Centro Latinoamericano de Demografía.

Ordorica, Manuel (1991), "Ajuste de una función expologística a la evolución de la población total de México, 1930-1985", *Estudios Demográficos y Urbanos*, vol. 5, núm. 3 (15), pp. 373-386.